

Mathematik für Information und Kommunikation am Beispiel des Huffman Algorithmus

Laut Heinrich Winter ist eines der allgemeinen Ziele des Mathematikunterrichts, „Erscheinungen der Welt um uns, die uns alle angehen oder angehen sollten, aus Natur, Gesellschaft und Kultur, in einer spezifischen Art wahrzunehmen und zu verstehen.“ Schüler sollen „exemplarisch Mathematisierung in Technik und Naturwissenschaften erleben“[1]. Um dieser Aufgabe gerecht zu werden, gehört dazu heute auch ein Verständnis von modernen Werkzeugen der Informations- und Kommunikationstechnik, die uns ja tagtäglich begegnen.

Am Beispiel des Huffman-Algorithmus soll hier aufgezeigt werden, wie in unserem Alltag Mathematik direkte Anwendung findet. Dieser Algorithmus wird bei vielen Kompressionsformaten eingesetzt, denen wir täglich begegnen. Dies ist z.B. der Fall beim ZIP-Format, JPEG¹-Format für Bilder, MPEG²-Format für Video, MP3³-Format für Audiodateien oder dem Telefax-Format, wobei der Huffman-Algorithmus als Teilschritt in einem komplexen und multiplen Kompressionsverfahren eingesetzt wird. Das Verfahren wurde 1952 von David Huffman (1925-1999) entwickelt und ist seither aufgrund seiner Einfachheit und Effektivität sehr beliebt [2].

Bildungsrelevante Aspekte des Huffman-Algorithmus

- Als ein Verfahren, das uns tagtäglich in unserer Umwelt begegnet, trägt es zum besseren Verständnis der „modernen Welt“ bei und dient damit der Umwelterschließung.
- Das Komprimierungsverfahren ist sehr elementar und kann mit einfachen Hilfsmitteln und ohne Computerkenntnisse nur mit Papier und Bleistift selbstständig durchgeführt werden.
- An diesem Verfahren lassen sich grundsätzliche algorithmische Methoden aufzeigen und algorithmisches Denken schulen.
- Fundamentale Prinzipien von Komprimierungsverfahren wie Optimierung, Häufigkeitsanalyse,... werden thematisiert.
- Der Huffman-Algorithmus arbeitet mit Baumstrukturen, die eines der wesentlichen Hilfsmittel der diskreten Mathematik sind. An diesem Beispiel können elementare Datenstrukturen in verständlicher Weise in der Schule dargestellt werden.

¹ Kompressionsstandard der **J**oint **P**icture **E**xpert **G**roup

² Kompressionsstandard der **M**otion **P**icture **E**xpert **G**roup

³ fällt unter den MPEG-Standard. Abkürzung für MPEG I Layer 3

- Der Huffman-Algorithmus ist eines der wenigen schulischen Beispiele für einen Algorithmus, der nicht eine Zahl als Endergebnis liefert, sondern einen (Code-) Baum. Das Ergebnis des Huffman-Algorithmus ist nicht notwendigerweise eindeutig, es können durchaus mehrere gleichwertige Lösungen existieren. Damit hebt sich das Verfahren deutlich von den gängigen rechnerischen Schulalgorithmen ab.
- Das Verfahren zeichnet sich durch eine hohe Beziehungshaltigkeit im Sinne Freudenthals aus. So kann beispielsweise über den Begriff der Entropie Zusammenhänge zu anderen Fächern hergestellt werden.

Die Grundidee des Huffman-Algorithmus

Bei vielen computerbasierten Codes wie beispielsweise dem ASCII⁴-Code werden jedem zu codierenden Zeichen die gleiche Anzahl an Binärzeichen zugeordnet. So benötigt man z.B. für das Wort „ABRAKADABRA“ $11 \cdot 8 \text{ Bit} = 88 \text{ Bit}$. Dies ist im Hinblick auf den Speicherplatzbedarf keine sehr effektive Codierung.

Effektiver wäre es, wenn man für häufig vorkommende Zeichen einen kürzeren Code vergäbe und selten auftretende Zeichen mit längeren Codewörtern codierte. Genau diese Idee wird beim Huffman-Algorithmus konsequent umgesetzt.

Diese Idee ist vergleichbar mit folgender Organisation eines Bücherregals: Häufig benötigte Bücher stellt man in greifbare Nähe (Augenhöhe), während man selten benötigte Bücher eher weiter oben oder unten verstaut. Auf Dauer spart man beim Holen der häufig benötigten Bücher sehr viel Aufwand ein. Dafür nimmt man es gerne in Kauf, selten genutzte Bücher aufwändiger (z.B. mit einer Leiter) zu holen.

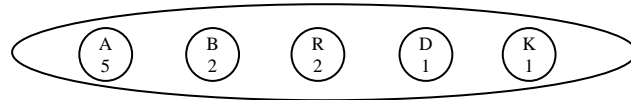
Ein anderes Beispiel für eine solche Codierung ist der Morse-Code, der von dem amerikanischen Malers und Erfinders Samuel Morse (1791-1872) entwickelt wurde. Zeichen werden hier mit Kombinationen von Punkten und Strichen codiert (damit meint man kurze oder längere Stromstöße). Die Anzahl der Punkte und Striche für jedes einzelne Zeichen variiert, je nach dem, mit welcher Häufigkeit es in der englischen Sprache auftritt. So wird beispielsweise das häufig vorkommende „e“ lediglich mit „.“, also einem kurzen Stromstoß codiert und der seltenere Buchstabe „q“ mit „-.-.-“ (lang lang kurz lang) codiert.

Der Huffman-Algorithmus exemplarisch an einem Beispiel

Ziel der Huffman-Codierung ist es, jedem im Text vorkommenden Zeichen einen Binärcode zuzuweisen. Dies soll an dem Wort „ABRAKADABRA“ exemplarisch gezeigt werden.

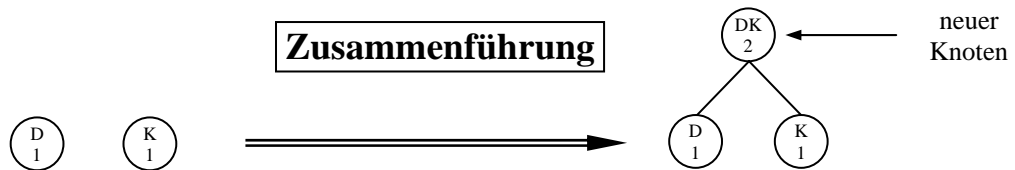
⁴ American Standard Code for Information and Interchange

Die Basis für den Algorithmus ist die Häufigkeit der zu codierenden Zeichen im Text. Deshalb ermittelt man zunächst, die Häufigkeit jedes Zeichen im Text. Der Huffman-Algorithmus arbeitet intensiv mit Bäumen. Zum Start des Algorithmus notiert man jeden Buchstaben mit der zugehörigen Häufigkeit in einem Knoten, den man allgemein gesprochen, auch als (Trivial-) Baum bezeichnen kann.

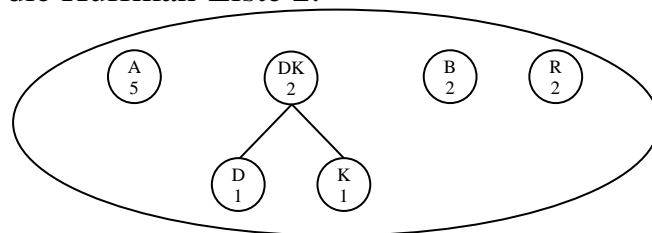


Diese Zusammenstellung sei als Huffman-Liste 1 bezeichnet.

Für den nächsten Schritt sucht man aus der Liste die zwei Knoten mit den geringsten Häufigkeiten heraus. In unserem Beispiel sind das die Knoten „D“ und das „K“. Diese führt man zu einem kleinen Baum, bestehend aus einem neuen Knoten zusammen, von dem jeweils zu den ausgewählten Blättern Kanten ausgehen. Der Knoten erhält den Namen der beiden Blätter und als Häufigkeit die Summe der Häufigkeiten der Blätter. Welcher der beiden Knoten dabei zum linken bzw. rechten Nachfolger gewählt wird, spielt für die Effektivität des Verfahrens keine Rolle. Dieser Schritt sei als die *Zusammenführung* der Knoten „D“ und „K“ bezeichnet.



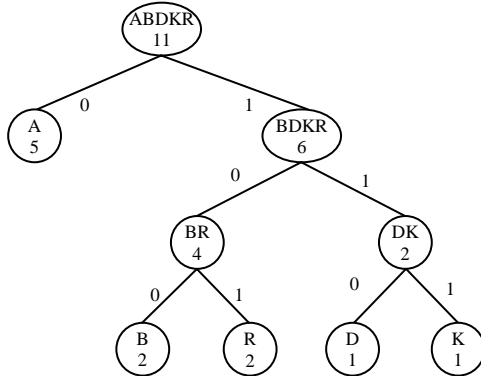
Anschließend wird der durch die Zusammenführung entstandene Baum in die Huffman-Liste einsortiert und ersetzt die zusammengeführten Blätter. Man erhält so die Huffman-Liste 2.



Im Folgenden führt man den Schritt der Zusammenführung von je zwei Bäumen mit der geringsten Häufigkeit solange aus, bis die Liste nur noch einen Baum enthält. Aus diesem Baum gewinnt man den Codebaum, wenn z.B. alle links abzweigenden Kanten mit „0“ und alle nach rechts abzweigenden Kanten mit „1“ beschriftet werden. Ein dazu entwickeltes Experimentiersystem ist unter [3] zu finden.

Mit Hilfe der Codetabelle erhält man für den Text „ABRAKADABRA“ die Codierung „01001010111011001001010“.

Codebaum



Codetabelle

Buchstaben	Binärkode
A	0
B	100
D	110
K	111
R	101

Mit nur 23 Binärzeichen ist dieser Code deutlich kürzer als der ASCII-Code. Typischerweise erreicht man durch das Verfahren eine Kompression auf ca. 2/3 der ursprünglichen Länge einer Textdatei [4].

Eigenschaften des Huffman-Codes

Der Huffman-Code ist ein Präfix-Code, d.h. kein Codewort eines Zeichens ist vollständig im Anfang eines anderen Codeworts enthalten. Damit wird das Decodieren eines Binärcodes sehr einfach, denn man startet bei der Wurzel des Baums und folgt den Binärzeichen des Codes, bis man ein Blatt erreicht hat. Ohne ein Trennzeichen können dann die weiteren Binärzeichen decodiert werden. In obigem Beispiel gelangt man mit dem ersten Zeichen „0“ zum A und mit den folgenden Zeichen „100“ zum „D“. Ein besonders eingängiges Beispiel für einen Präfix-Code ist das (Festnetz-)Telefonnummernsystem.

Bei der Erstellung des Huffman-Baums hat man oft eine gewisse Wahlfreiheit bei der Auswahl, welche Bäume man im nächsten Schritt zusammenführt, denn oft haben mehrere Bäume die gleiche geringste Häufigkeit. Man kann zeigen, dass dies keine Auswirkung auf die Effizienz des Verfahrens hat. Der Huffman-Algorithmus liefert in jedem Fall den kürzest möglichen Präfix-Code.

Eine ausführliche Darstellung des Verfahrens und seiner Anwendung beim Telefax ist zur Veröffentlichung in der Zeitschrift der Mathematikunterricht (Anfang 2006) vorgesehen. Dieser Beitrag entstand unter Mitwirkung von Herrn Christan Urff. Für eine anregende Beratung möchte ich mich gerne bei Herrn Prof. Dr. Ziegenbalg bedanken.

- [1] Heinrich Winter, „Mathematikunterricht und Allgemeinbildung“, In: Mitteilungen der Gesellschaft für Didaktik der Mathematik, H. 61/1995
- [2] David A. Huffman, "A Method for the Construction of Minimum-Redundancy Codes" Proc. IRE, pp. 1098-1101, September 1952
- [3] Internet-Portal zur Veranstaltung „Codierung und Kryptographie“ (Ziegenbalg): <http://www.ph-karlsruhe.de/~ziegenbalg/materialien-homepage-jzbg/cc-interaktiv/>
- [4] Uwe Schöning, „Algorithmik“, Spektrum Akad. Verl., 2001 S. 252